Quality of diagnostic teaching abilities in early education

van Oers, H.J.M.

published in
European Early Childhood Education Research Journal
1999

DOI (link to publisher)
10.1080/13502939985208391

document version
Publisher's PDF, also known as Version of record

European Early Childhood Education Research Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t776628938

Quality of diagnostic teaching abilities in early education

Bert van Oers[a]
[a] Free University Amsterdam, The Netherlands

To cite this Article van Oers, Bert(1999) 'Quality of diagnostic teaching abilities in early education', European Early Childhood Education Research Journal, 7: 2, 39 — 51
To link to this Article: DOI: 10.1080/13502939985208391
URL: http://dx.doi.org/10.1080/13502939985208391

European Early Childhood Education Research Journal
Vol. 7, No. 2, 1999

Quality of Diagnostic Teaching Abilities in Early Education

BERT VAN OERS

Free University Amsterdam
The Netherlands

SUMMARY *In a longitudinal study of second year teachers in a Dutch primary school, an analysis was performed of the teachers' diagnostic teaching abilities in the domains of literacy and numeracy. Their pupils' literacy development was assessed using standardised tests, and the results of these tests compared with the teachers' criterion-referenced judgment of their pupils' literacy development. The teachers' judgments proved to be valid and reliable.*

RÉSUMÉ *Une analyse de la capacité des instituteurs à établir un diagnostic dans le domaine de la lecture et du calcul a été effectuée au cours d'une étude longitudinale menée auprès d'instituteurs de la deuxième classe d'une école primaire néerlandaise. Le niveau des élèves en lecture a été évalué au moyen de tests standardisés et les résultats à ces tests ont été comparés aux jugements personnels des instituteurs. Ces jugements se sont révélés valides et fiables.*

ZUSAMMENFASSUNG *In einer Längsschnittstudie in der zweiten Klasse einer niederländischen Grundschule wurde eine Analyse über die Qualität der diagnostischen Lehrmethoden auf den Gebieten des Lesens und Rechnens durchgeführt. Die Entwicklung der Lesekompetenz wurde mit einem standardisierten Test ermittelt. Die Ergebnisse wurden mit den Beurteilungen von Lehrkräften aufgrund ihrer Erfahrungen im Alltag verglichen. Die Beurteilungen der Lehrer erwiesen sich als zuverlässig.*

RESUMEN *En un estudio longitudinal realizado en el segundo curso de una escuela primaria holandesa, se evalúa la aptitud del profesorado para enseñar lectura y aritmética según el método diagnostico. Por otra parte se utilizo un test normalizado para determinar el nivel de lectura de los niños. Estos datos se compararon con las calificaciones dadas por los maestros, basándose en sus experiencias diarias. Se concluyo que las calificaciones de los maestros eran validas y fidedignas.*

Keywords: Diagnostic teaching; Play; Literacy; Teacher report; Testing.

Appropriate teaching

The appropriateness of teaching largely depends on the teachers' abilities to construct an adequate curriculum with the help of the various resources that are available and relevant at a particular moment. In their critical evaluation of the technocratic, 'remote control', perspective on curriculum development Pinar et al (1995) quite rightly emphasise:

> "Where the curriculum can be constructed now is the 'lived space' of the classroom, in the lived experience of students and teachers" (Pinar et. al, 1995, p. 860).

It is obvious that the teachers' ability to establish the nature of this 'lived space' is an essential element in the success of this style of teaching and curriculum development. This appropriate teaching style that intends to construct the curriculum interactively in the classroom, taking into account both main educational aims and pupils' individual qualities (abilities, interests, emotions), is here called *diagnostic teaching.*

In my research programme I am investigating teaching and learning processes in the context of a play based curriculum for primary school. One of the things that we have to find out is how teachers can innovate their style of classroom activities in accordance with this diagnostic teaching approach. In order to gain some insight in the diagnostic teaching abilities of teachers involved in this approach, we decided to design a longitudinal study (starting in August 1997). As the play based curriculum is currently just implemented in the early grades of some of the Dutch primary schools, the study focuses only on the early grades (2 - 4, ages 5 - 8 years old). We are currently gathering data concerning teachers' way of observing children's activities, learning outcomes, and developmental progression in the domains of early literacy and numeracy in the grades 2 through 4 in one Dutch primary school.

In this article I want to describe some of our first findings from this research, focusing in particular on early literacy development in grade 2. In the first stage of the study, I wanted to find out,

a) what observational categories are currently being used by these teachers with regard to literacy development?

b) how do the teachers' descriptions of the pupils' development with regard to literacy development change as a result of assistance focusing on the basic strategy of observing and registering pupils' development?

c) is there any correspondence between the criterion referenced assessment of the pupils development by the teacher, and the more formal (norm-referenced) test-wise assessments in the domains of early literacy (validation of the observation strategy of the teachers)?

The educational approach

The teachers involved in our research are all employed at a school in the Netherlands that has adopted the concept of 'developmental education' as a basis for teaching. The main tenets of this approach are drawn from the sociocultural (Vygotskian) approach to human development. This theoretical approach has been elaborated into a curriculum strategy, embodied in a number of resources (documents, video-tapes, in-service training), intending to assist teachers in the attempt to realise appropriate teaching for all pupils in primary schools (aged 4 - 12). These efforts resulted in a play-based curriculum called 'Basic Development'[1] for the early grades of primary school (ages 4-8). Play is seen as the fundamental context for learning for children of this age (Vygotsky, 1978; El'konin, 1972); in the context of play teachers should identify (or create) meaningful teaching opportunities (van Oers, in press; Janssen-Vos & van Oers, 1998).

It is impossible to give a detailed description of the curriculum strategy here, but a few elements have to be mentioned as they are considered especially important for the literacy development of the pupils. Firstly, literacy learning processes are promoted and improved in the context of the play activities of the children. This is practised by the teachers working according to this developmental education ap-

proach. For the promotion of learning processes in the context of play it is considered of major importance that the teacher participates in the play activity in order to give assistance or stimulation/encouragement where needed. A second tenet of this approach to literacy education is, that reading and writing have to be considered as closely related processes. Pupils in our schools learn to read from a writer's perspective. The activity of reading another writer's texts emerges from this activity of writing and reading your own texts (Knijpstra et. al., 1997). Finally, as writing is seen as a specific representational activity, much attention is given to the early representational activities of the children (drawing, making diagrams, writing to drawings etc. in the context of a variety of play activities of the children (van Oers, 1994; 1997).

As to the teaching activity, two major elements of this approach consist in the teachers' ability to *observe* the pupils meticulously in order to gather information about the developmental process of a pupil involved, and to *design* a new development-promoting activity for these pupils accordingly. Spotting or creating productive teaching opportunities, related to both educational aims and the pupils' actual qualities is basically the core of this educational approach (van Oers et al., 1996).

As to the educational aims articulated by the curriculum strategy, there is an inventory of developmental domains and aims, that have to be covered by the teachers in the early grades. The course of development of the individual pupils, however, is not uniformly prescribed, but depends on the needs that the teacher observes in the individual pupils or the need that the teacher can provoke corresponding with the system of educational aims. The important logistic rules for the teacher to make choices from the inventory are the following:

a) first guarantee the pupils well-being, and take care that the pupil feels safe and personally engaged;

b) broad developmental aims (communicating, reflecting, problem solving, making symbols etc.) should precede more specific technical skills (e.g. reading and writing techniques, spelling, counting, etc.); broad developmental aims are to be defined in terms of involvement in socio-cognitive activities, without putting much emphasis on the technical mastery of all the skills required in this activity. The idea is that we first should engage pupils in an activity, before bringing it to perfection by improving the actions included. Basically, specific skills always have to be developed from needs that emerge in the context of the more general (unspecified) activities.

Related to this curriculum strategy an observational strategy has been developed that identifies the main observational categories with respect to the broad developmental characteristics and the specific skills of (e.g.) reading, writing, numeracy. This observation strategy is documented and described in a manual (Janssen-Vos, Pompert & Schiferli)[2] . In accordance with the theoretical framework of the Developmental Education concept, a fundamental assumption of the observation strategy is that actions to be assessed have to be embedded in activities that are meaningful for the child. It is assumed that the isolation of actions from their activity contexts in order to estimate their qualities (as often happens in paper-and-pencil tests), actually deprives these actions of one of their basic resources (meaning, context), and will impair the performance of the pupils. This theoretical starting point is supported by the research of Donaldson and her colleagues, who have shown that tasks that make human sense for children offer better opportunities for children to show their actual cognitive capacities (see for example Donaldson, 1978; Hughes, 1987).

Hence, teachers in schools working from this concept of Developmental Education closely observe children in the context of meaningful (authentic) activities (i.e. play for the early grades of primary school). These observations are registered in an individual diary for each child, producing assessments of authenticated actions. The basic intention of this diary keeping is the construction of an overall record of each

pupil's development on a day-by-day basis. These diary recordings are put together with the main products of the children's activities (e.g. written texts, drawings, worksheets) into a portfolio that can be consulted by the teacher on a later occasion.

Problems in the appropriation of developmental teaching strategies

However, the practical experiences of teachers working within this perspective demonstrated that the development of abilities to observe and register the pupils' development was a complicated process, as there were no instruments available that strictly regulated or exactly prescribed the teachers' diagnostic actions. Teachers sometimes felt uncertain about how to apply the curriculum strategy and the observational strategy. In a small-scale intervention study it could, however, also be demonstrated that systematic assistance of teachers in writing logbooks about their planning of play and teaching, and in writing diaries about the individual children's development strengthened the self-confidence and observational quality of the teachers when they were working in accordance with this approach (Holla, 1997; see also van Oers & Holla, 1997). In general, these findings are consistent with what is found in the literature on teacher training: systematic assistance of teachers, close to their daily classroom work, contributes to the quality of teaching.

A study of teachers' observations

Context of the study

As we are currently still working on detailed analyses of the data that we gathered during the first year of this study, we can only present here a part of the information that we can expect to obtain. The final intention of the study is to follow the present grade 2 group (1997-1998) over three years. The performances of the pupils in the grades 3 and 4 were tested in order to be used as controls and as base-level performances for the future performances of the pupils of the present grade two when they arrive in the grades 3 and 4.

The research is being conducted at a school in a small town in the North-Holland region of the Netherlands. Nine teachers at this school from the lower grades (2-4; pupils ages from 5 - 8) were involved in this study. In order to follow the individual development of the pupils, we gathered data regarding the pupils' performances in the domain of literacy along two different lines:

1) Pupils in the grades 2 - 4 were individually tested with respect to functional literacy level by a teacher not employed at this school (in order to maximise independent assessment). These tests were taken three times during the first year (October '97, February '98, June '98) for literacy.

2) Teachers from the second grade wrote brief developmental reports about their pupils on the basis of their diaries, focusing on the development of the pupils in the domains of literacy and numeracy. These reports were written especially for this investigation. They did this three times during the school year (October '97, February '98, June '98). The teachers were instructed to follow the observation guidelines from the Observation manual. They were encouraged to include all the information they considered to be important for a valid description of the pupils' development in the domains of literacy and numeracy.

Observation and test instruments

The tests used for literacy development were a Dutch standardised test for individual reading development (AVI; Visser et al., 1996). This AVI-test is widely used and validated in the Netherlands) for the groups 3 through 6. The test is based on observations of children reading aloud a standard text (from a levelled series of booklets). The test intends to establish individual reading achievement of pupils, taking into account the time needed for reading a text and the number of reading errors made. For the youngest pupils (grade 2) a Dutch translation of Clay's 'Concept about Print' was used (Clay, 1993a). Clay's view on literacy development appears to be close to our Developmental Education approach to literacy (Clay, 1993b; Clay & Cazden, 1990). Clay's test is a diagnostic test for the assessment of young children's ideas about text, reading, and writing. The procedure of the text is based on a conversation of the teacher with an individual pupil about a book that they read collaboratively. Meanwhile the teacher asks the child different questions about where to start reading, in what direction one must read, invites the pupil to read with the teacher by pointing to the words that are being read etc. On each subsequent testing through the year another book of a comparative level is taken. The test contains 23 items and has been validated for different countries over the world, obtaining good reliability scores (see Clay, 1993a, p. 50). In our investigation similar results were obtained (alpha = 0,78 - 0,80; test-retest (Spearman R) correlation is .73, which is significant at the p= 0.01 level). These results indicate that the Clay test gives a reliable measure of these young children's literacy abilities. These tests were administered on 38 pupils (mean age: 72 months in March 1998).

During the first year of the investigation the teachers of grade 2 pupils wrote developmental reports about the children, especially with regard to general developmental qualities (play quality, social relations, general attitude, interests), to literacy development and to numeracy development. The teachers used the observation manual as a guideline, but they were free to compile their reports as they thought would be relevant and necessary. The reports were made in the same period as the tests were taken. The teachers did not know about the test results, but wrote their reports on the basis of their own classroom observations. The length of the teacher reports were rather similar (varying from 1/2 to 1; size A4). For the investigation we analysed the teacher reports of 38 children.

Analysis of the teachers' reports

We conducted two kinds of analyses on the teachers' reports. First we did a content analysis to get an idea of the kind of categories (with regard to both literacy and numeracy) teachers are using when they are observing children during play activities. Later we transformed the teachers' reports into an ordinal score in order to find out statistically if there was any correspondence between the teachers's criterion referenced assessments and the more formal testing with the help of the Clay test.

The results of this first content analysis (October 1997) regarding literacy can be summarised in the following tables.

Of course, Table 1 is just a gross indication of the categories used by the teachers when they are observing pupils. It is clear that these teachers working with these young children do not frequently use categories related to a decoding view on the reading process. Instead, their observations are focused - as could be expected from their approach - on categories related to viewing literacy activities as participating in literate activities (communicating, writing, listening to stories, telling stories).

When we look at how the teachers use their categories in their observations, it turns out that teachers tend to be very consistent: when they use to observe a particular

TABLE 1: Categorie used by teachers (October 1997)

Category ? teacher<	A	B	C	D	E	F
participation in literate activities	x	x	x	x	x	x
making a story to a drawing	x				x	x
retelling a story					x	x
reproduce a word (writing, magnetic letterboard etc.)	x	x			x	x
reading/writing thematic words	x	x			x	x
reading/writing name	x	x		x		x
recognition of letters	x	x	x	x	x	x
interest in books, texts	x	x	x	x	x	x
correct writing direction	x	x				
decoding		x				
word analysis & synthesis	x	x			x	x
symbol-phoneme connection		x				
use of 'book language'	x	x				x
oral language	x	x	x			x
pencil grip		x				
Number of categories used by one teacher	11	13	4	4	8	11

category, they take that category into account in the observation/description of every child. So we may conclude that the within-teacher consistency is very high. However, there are only three categories used by all teachers, while half the number of categories is used by just 1, 2, or 3 teachers. Hence we may conclude that the between-teacher consistency (or the homogeneity of the team as to the observation of pupils) is not very high.

We must be very careful with the interpretation of this table as an explanation of the teachers' diagnostic teaching ability. This table only shows those categories that were considered to be relevant for these teachers with these pupils at this particular moment. We may expect that categories change when pupils grow older and that the teachers assume new qualities to be relevant for assessing the development of these older pupils. The most important fact is that the teachers obviously look at those qualities that can be considered to be highly relevant from the Developmental Educational point of view. This strengthens the assumption of the content validity of their observational strategy. Nevertheless, I was a little concerned about the low homogeneity of the team as to observational vocabulary. Hence we gave some feedback in a meeting with all the teachers involved, discussing the observational categories in relationship with the observation strategy as was described in the HOREB-manual. This meeting was organised by an external counsellor, who is familiar with the educational concept of the school and with the observation strategy used.

In the next round of reports (March 1998) we repeated the content analysis, and we obtained the following results.

TABLE 2: Categories used by teachers (March 1998)

Category **teacher<** **?**	A	B	C	D	E	F
participation in literate activities	x	x	x	x	x	x
making a story to a drawing	x			x	x	
retelling a story	x			x	x	x
reproduce a word (writing, magnetic letterboard etc.)	x	x	x	x	x	x
reading/writing thematic words	x	x				
reading/writing name				x		x
recognition of letters	x	x	x	x	x	x
interest in books, texts	x	x	x	x	x	x
correct writing direction	x				x	x
decoding		x			x	
word analysis & synthesis	x	x		x	x	x
symbol-phoneme connection						
use of 'book language'						x
oral language	x	x		x	x	x
pencil grip						
sentence construction	x	x				x
writing texts/booklet	x	x	x	x		x
read own self-written texts	x	x	x		x	x
independent reading of books	x	x	x	x	x	x
spelling					x	
operations not yet mastered	x	x	x	x	x	x
Number of categories used by one teacher	15	14	12	12	14	15

The comparison of these two tables shows that teachers have added some categories to their observational repertoire (see the bold faced categories in the table above) and left out others. This is not surprising as the new categories refer to higher levels of reading/language proficiency that may be expected with the growing age of the pupils.

In general we can see that the observational vocabulary of the teachers has become more homogeneous, both quantitatively (they use approximately the same number of categories now) and qualitatively (they employ more communal categories). This latter fact may be due to our general feedback on the first reports which may have stimulated the teachers to use the observation manual more explicitly as a resource.

In order to do a quantitative analysis I transformed the teachers' reports to an ordinal score indicating the level of literacy development as described by the teachers. It is very important that the teachers can identify pupils who need special support with their own observational strategy. Hence, I wanted to examine whether the ordering of pupils with the help of the teachers' diagnostic strategy somehow correlates with an ordering on a standardised scale. I constructed the following levels from the teachers reports:

0 Teacher contends that pupils shows *no or very little interest in written language;* pupil does not take part in conversation or seldom tells spontaneously a personal experience; will not volunteer for playing in the book/library/reading corner; teacher shows concern about pupils development; little or no mastery of specific reading skills (-). **Preliterate pupil.**

1 Teacher reports *some interest in written language* in the pupil, but only in the context of play activities; the pupils likes to listen to stories and is eager to tell from own experience; sometimes - but not very often - the pupils chooses to play in the library/book/reading corner of the classroom; pupils masters a few basic specific reading skills (for example letter recognition, a little word analysis, writes its own name etc); *many gaps* in the mastery of relevant reading/writing skills (-/+); **beginning reader.**

2 Teacher reports great interest in books and written and oral language; pupils expresses himself easily and often tries to write something both in and outside play situations; can read easy texts for himself; most basic reading skills are mastered, only a few skills are not yet well automatised/mastered; (++-). Child reads children's books easily and fluently. **Advanced reader.**

3 Proficient reading and writing of all kinds of written material (+++). **Proficient reader.**

Summarising the scores for the pupils on this scale (first assessment, Nov. '97) shows the following result:

TABLE 3: Reading levels of pupils (November 1997)

SCORE	N	%
0	5	13.2
1	16	42.1
2	15	39.5
3	-	-

The results on the different assessments on the Clay instrument can be summarised in the following table:

TABLE 4: Scores on the Clay test ('Concept about print')

	N	Range	Min.	Max.	Mean	St. dev	Var.
Clay 1	36	18	1	19	10.13	3.39	11.50
Clay 2	36	10	9	19	14.35	2.83	8.01
Clay 3	32	13	8	21	16.51	3.05	9.32

As can be read from this table, the mean score of the test is gradually rising (the difference is statistically significant at p= 0.01; two-tailed t-test for related samples), while the variance in the group is descending a little. The literacy level of the group is improved, while the heterogeneity is slightly reduced.

Calculating a nonparametric correlation between these test results and the results on the Clay tests shows significant correlations at the 0.01 level (Kendall's tau: .48; Spearman Rho: .593 for the first assessment; .48 and .56 respectively for the second assessment; .48 and .55 for the third). This means that on both occasions the ordering of pupils according to both tests is more or less the same: the teachers' own assessments on the basis of day-by-day observations in the context of meaningful play activities of the children can identify the at-risk pupils (with respect to the literacy development) as well as a more formalised test. The judgments of the teachers turn out to be stable, as the correlations between the first assessment and the second is also significant at the 0.01 level).

In order to get a more detailed view on the teachers' ability to identify pupils who may need special attention, we scrutinised the data with respect to this aspect. The following summary could be made for the first assessment (November 1997):

TABLE 5: Teachers' assessments compared to scores on the Clay test (November 1997)

	Grouped Clay levels			
		0 - 10	11 - 18	19 >
levels based on teachers' judgment	**0**	4	1	
	1	14	2	
	2	3	11	1

N = 36

This table shows that most pupils in need of more attention for literacy development (according to the Clay-scale) were also identified by the teachers as at-risk pupils. Only

one pupils is diagnosed as 'problematic' while he has an average score on the Clay-scale. This pupil may get more attention than he may need considering his test score on the Clay instrument. Very worrying, however, is the assessment of 3 pupils as demonstrating a good literacy development according to the teachers, while their scores on the Clay-scale are low. It seems that their capacities have been overestimated by the teachers. These pupils may get less attention than they may actually need.

In the second round of teacher assessments (March 1998), it turned out that teachers were a little more conservative with their judgments of the pupils literacy abilities (probably the pupils have shown less development than the teachers expected in the beginning). The table is then as follows:

TABLE 6: Teachers' assessments compared to scores on the Clay test (March 1998)

	Grouped Clay levels			
		0 - 10	11 - 18	19 >
levels based on teachers' judgment	**0**	2	3	
	1	1	19	
	2		10	1

N = 36

This table shows that the teachers were more cautious in their estimations of the pupils abilities. No pupils have been identified as having a good literacy development, while their scores on the standardised test turned out to be low. Apparently, the teachers had observed the pupils during a longer time and were able to establish a better (more realistic?) estimate of the pupils' literacy abilities. It must be repeated here that the teachers were not informed about the pupils scores on the Clay or any other test. Their view is built upon the information that they gather from their day-by-day classroom observations of children.

By the end of the school year (June 1998) they have been able to get to know the pupils better and can give even better founded assessments. The results then look like Table 7. The table speaks for itself. It is clear that the assessments of the teachers are stable over the year. There are no at risk pupils that haven't been noticed by the teacher, while the pupils with high literacy abilities (according to the Clay-instrument) have been identified by the teacher as well. Again we may conclude that the teachers' daily involvement with the children, provides ample opportunities for observing the children and assessing their literacy qualities in a reliable way.

Discussion and conclusions

From our present data and analyses we are inclined to conclude that teachers can assess pupils literacy abilities on the basis of their day-by-day classroom observations and registrations, provided they get adequate help and can work from a clear theoreti-

TABLE 7: Teachers' assessment compared to scores on the Clay test (June 1998)

	Grouped Clay levels			
		0 - 10	11 - 18	19 >
levels based on teachers' judgment	**0**	1	1	
	1		16	2
	2		6	6

N = 36 (during the school year 3 pupils have left the school; one pupil was absent)

cal view that helps them to identify the qualities that have to be recorded in order to construct a reliable and valid view of pupils literacy abilities. The importance of this finding is to be seen in the legitimacy that we obtain for constructing judgments of pupils on the basis of continuous observations in situations that are meaningful for the pupils themselves. Obviously it is not (always) necessary to test a pupil's special abilities with a formal test, when we have already observed that ability in the living and meaningful activity of the child.

Of course, it is necessary then that teachers are well informed about the nature of observing children and recording relevant qualities of the child. In our sample of teachers it also turned out that the teachers were not equally capable of doing such observations and/or recordings. Sometimes teachers are uncertain about their diagnostic teaching abilities and need support in order to appropriate this style of teaching.

On the basis of the data discussed above, it is possible to conclude that the mean level of literacy activity of the pupils is significantly raised over the year and that the teachers become more and more equipped to assess the children's development independently on the basis of their daily observations. Although the Clay-test was taken three times during the year, the improvement cannot be a simple result of practice on the test, since at every test occasion another booklet was taken to read with the children. The books can be considered to have the same level of difficulty. Three of the four missing values were slightly (less than one standard deviation) below or above the means on the first and second test occasion and were all considered 'beginning readers' by the teacher (the fourth was just absent). This might have contributed to the rise of the mean in the group. However, the mean score of the group was already ascending between the first and second occasion (when these pupils were still in the group). The difference between test occasion one and two is even more than between two and three. The difference between occasion one and two is statistically significant at the $p = 0.01$ level. We are inclined to conclude that the improvement of the literacy scores reflect a genuine tendency in this group. The ability of the teachers to tune in more and more precisely on the exact level of the pupils is probably one of the driving forces behind that effect.

Although the results of this research are promising until now, there is much left for further questioning. That is why we will continue to follow the developments of these pupils in this group for the next few years.

A final remark relates to our research strategy. Recent discussions on achievement testing demonstrate the tendency for more criterion referenced testing instead of

(or in addition) to norm-referenced testing (Hambleton & Sireci, 1997). Our teachers' way of diagnosing children in the course of the teaching process is strongly inspired by a criterion-referenced approach. We have no reason now to refuse all kinds of norm-referenced testing. At least we need the norm-referenced test in order to validate our criterion-referenced observation strategies. The situation may change, however, when we have more information about the validity of the criterion-referenced testing based on classroom observations by the teacher. But even then, the quality of the outcomes largely depends on the diagnostic teaching abilities of the teacher. With regard to this matter, we carefully tried to avoid the expression of 'authentic' assessment, as this may wrongly suggest that this kind of assessment is by definition better or more reliable due to its so-called authenticity. We can agree with Terwilliger's critical argument regarding this issue (Terwilliger, 1997). However, Terwilliger fails to recognise the need for assessment reform per se, i.e. the need for improved strategies for constructing reliable views of children's abilities. The focus should not be on authentic assessment per se, but on assessment in the context of authentic (meaningful) activities of children. Our research has demonstrated that steps towards such reliable and valid assessments can be made.

NOTES

[1] This curriculum ('Basisontwikkeling', in Dutch) is developed during the past decennia in The Dutch Non-denominational Pedagogical Institute (Utrecht), by Frea Janssen-Vos and her colleagues. It is now being implemented in a growing number of school in the Netherlands (Janssen-Vos, 1997). Deliberately, this curriculum is not a ready-made document telling the teachers on a day-by-day or hour-by hour basis what to do. The approach is better called a curriculum *strategy* (embodied in a series of books, videos, and articles), that intends to assist the teachers in developing an adequate curriculum for all pupils of the learning group.

[2] This manual is called HOREB, which is a Dutch abbreviation for 'Action oriented observation and registration in the Basic Development-programme').

REFERENCES

CLAY, M. M. (1993a) *Becoming Literate: The Construction of Inner Control* (Auckland: Heineman).

CLAY, M. M. (1993b) *An Observation Survey: Of Early Literacy Achievement* (Auckland: Heineman).

CLAY, M. M. & CAZDEN, C. B. (1990) A Vygotskian Interpretation of Reading Recovery, in: MOLL, L. (Ed.) *Vygotsky and Education. Instructional Implications and Applications of Sociohistorical Psychology* (Cambridge: Cambridge University Press, pp. 206-223).

DONALDSON, M. (1978)*Children's Minds* (London: Fontana).

EL'KONIN, D. B. (1972) Toward the problem of stages in the mental development of the child, *Soviet Psychology*, 10, 225-251.

HAMBLETON, R. K. & SIRECI, S. G. (1997) Future directions for norm-referenced and criterion-referenced achievement testing, *International Journal of Educational Research* (Special issue on: Educational testing and assessment: Lessons form the past, directions for the future; edited by L.W. Anderson), Vol. 27, nr 5, 379-394.

HOLLA, A. (1997 September) *How Teachers Learn to Work with the Concept of Developmental Education*. Paper for the 7th conference of the European Early Childhood Education Research Association, Munich.

HUGHES, M. (1987) *Children and Number. Difficulties in Learning Mathematics* (Oxford: Blackwell).

JANSSEN-VOS, F., POMPERT, B. & SCHIFERLI, T. (1997) *Het HOREB instrument* (Utrecht: APS).

JANSSEN-VOS, F. & VAN OERS, B. (1998) Basisontwikkeling: een Vygotskiaanse strategie voor ontwikkelingsstimulering, in: HARPMAN, A., VEENKER, H, & POL, G. (Eds.) *Praten, Denken, Doen* (Alphen aan den Rijn: Samsom, pp. 183-202).

KNIJPSTRA, H., POMPERT. B., SCHIFERLI, T. (1997) *Met Jou Kan ik Lezen en Schrijven* (Assen: van Gorcum).

PINAR, W. F., REYNOLDS, W. M., SLATTERY, P., & TAUBMAN, P. M. (1995) Understanding Curriculum (New York: P.Lang).

TERWILLIGER, J. (1997) Semantics, psychometrics, and assessment reform: a close look at 'authentic' assessments, *Educational Researcher*, Vol. 26, nr. 8, (November), 24-27.

VAN OERS, B. (1994) Semiotic activity of young children in play: the construction and use of schematic representations, *European Early Childhood Education Research Journal*, Vol. 2, no 1, 19-34.

VAN OERS, B. (1997) The narrative nature of young children's iconic representations: some evidence and implications, *International Journal of Early Years Education*, vol. 5, nr 3, 237-246.

VAN OERS, B., JANSSEN-VOS, F., POMPERT, B. & SCHIFERLI, T. (1996) La pédagogie: une activité conjointe, in: RAYNA, S., DELAU, M. & LAEVERS, F. (Eds.) *L'education Préscolaire: Quels Objectifs Pédagogiques?* (Paris: Nathan, pp. 235-255).

VAN OERS, B. & HOLLA, A. (1997) *The Milestone: An Implementation of Developmental Education. Case-study for the European Observatory of Innovations in Education and Training* (Amsterdam/Paris: Free University/INRP).

VISSER, J., VAN LAARHOVEN, A., & TER BEEK, A. (1996) *AVI-toetspakket (3rd edition)* ('s Hertogenbosch: KPC).

VYGOTSKY, L. S. (1978) *Mind in Society* (Cambridge: Harvard UP).

Correspondence about this paper should be addressed to:

Bert van Oers
Department of Education
Free University Amsterdam
van der Boechorststraat 1,
1081 BT Amsterdam
The Netherlands